DOCUMENT RESUME

ED 435 084 CS 013 668

TITLE	High-Stakes Assessments in Reading. A Position Statement of the International Reading Association.
INSTITUTION	International Reading Association, Newark, DE.
PUB DATE	1999-00-00
NOTE	19p.
AVAILABLE FROM	International Reading Association, 800 Barksdale Road, P.O. Box 8139, Newark, DE 19714-8139 (single brochure copies free on request when accompanied by stamped, self-addressed envelope). Web site: <http://www.reading.org>.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; Position Papers; *Reading; *Reading Achievement; *Reading Skills; *Reading Tests; *Student Evaluation
IDENTIFIERS	*High Stakes Tests; International Reading Association

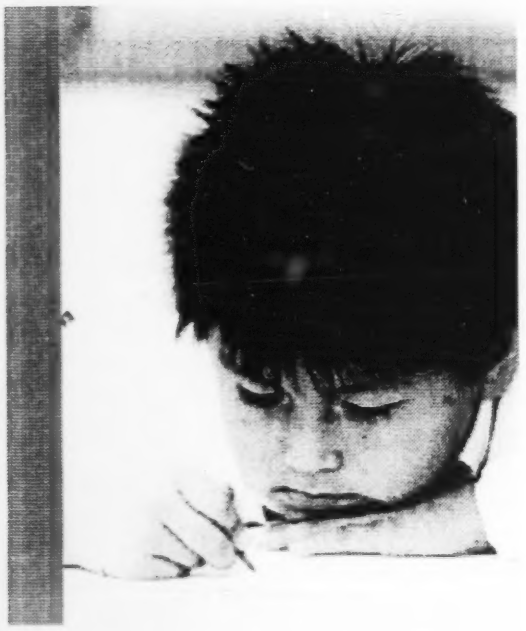
ABSTRACT

High-stakes testing means that one test is used to make important decisions about students, teachers, and schools. This paper comes out against high-stakes testing, taking the position that the central concern is that testing has become a means of controlling instruction as opposed to a way of gathering information to help students become better readers. The paper, in the form of a question-and-answer dialogue, aims to guide educators who must use tests as a key element in the information base used to make decisions about the progress of individual children and the quality of instructional programs. The paper poses seven questions and provides focused discussions on each question. It also offers recommendations for teachers; researchers; parents, parent groups, and child advocacy groups; and policymakers. As a frame for these recommendations, the paper stresses two points: (1) accountability is a necessary part of education; and (2) the intent is not to blame policymakers for the current dilemma with high-stakes testing. (NKA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 084

High-Stakes Assessments in READING



A Position Statement of the International Reading Association

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. M. Irwin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CS 013 668

The Board of Directors of the International Reading Association is opposed to high-stakes testing. High-stakes testing means that one test is used to make important decisions about students, teachers, and schools. In a high-stakes testing situation, if students score high on a single test they could be placed in honors classes or a gifted program. On the other hand, if students score low on a high-stakes test, it could mean that they will be rejected by a particular college, and it could affect their teacher's salary and the rating of the school district as compared with others where the same test was given.

In the United States in recent years there has been an increase in policy makers' and educators' reliance on high-stakes testing in which single test scores are used to make important educational decisions. The International Reading Association is deeply concerned about this trend. The Board of Directors offers this position statement as a call for the evaluation of the impact of current types and levels of testing on teaching quality, student motivation, educational policy making, and the public's perception of the quality of schooling. Our central concern is that testing has become a means of controlling instruction as opposed to a way of gathering information to help students become better readers. To guide educators who must use tests as a key element in the information base used to make decisions about the progress of individual children and the quality of instructional programs, we offer this position in the form of a question and answer dialogue. This format is intended to ensure that important conceptual, practical, and ethical issues are considered by those responsible for designing and implementing testing programs.

What does the term *high-stakes testing* mean?

High-stakes testing means that the consequences for good (high) or poor (low) performance on a test are substantial. In other words, some very important decisions, such as promotion or retention, entrance into an educational institution, teacher salary, or a school district's autonomy depend on a single test score.

High-stakes tests have been a part of education for some time. Perhaps the most conspicuous form of high-stakes testing, historically speaking, was in the British educational system. National exams in England and in other countries that adopted the British system separated students into different educational tracks. In the United States, tests such as the Medical College Admission Test and Law School Admission Test, as well as professional certification examinations (for example, state bar examinations, medical board examinations, state teacher examinations) all represent high-stakes tests.

The meaning of high stakes can be confusing at times. Tests that have no specific decision tied to them can become high stakes to teachers and school administrators when they must face public pressure after scores are made public. In other cases, a low-stakes state test can be transformed into a high-stakes test at a school district level if a local school board decides to make educational or personnel decisions based on the test results.

Why are we concerned with high-stakes testing?

Although high-stakes testing has been and probably will continue to be part of the educational landscape, there has been an increase in such testing in recent years, particularly at the state level. More children are being tested at younger ages, and states and local school districts are using these tests to make a greater variety of important decisions than ever before. Increased frustration with lack of achievement has led to a greater reliance on testing. In response to these frustrations many states have adopted educational standards and assessments of those standards. The logic is that tests of standards accompanied by a reward and penalty structure will improve children's achievement. In too many cases the assessment is a single multiple-choice test, which would be considered high stakes and would not yield enough information to make an important instructional decision.

Is testing an important part of good educational design?

Yes, testing students' skills and knowledge is certainly an important part of education, but it is only one type of educational assessment. Assessment involves the systematic and purposeful collection of data to inform actions. From the viewpoint of educators, the primary purpose of assessment is to help students by providing information about how instruction can be improved. Assessment has an important role to play in decision making beyond the classroom level, however. Administrators, school board members, policy makers, and parents make significant decisions that impact students. The needs of many audiences must be considered in building a quality assessment plan.

Testing is a form of assessment that involves the systematic sampling of behavior under controlled conditions. Testing can provide quick reliable data on student performance. Single tests might be used to make decisions that do not have major long-term consequences, or used to supplement other forms of assessment such as focused interviews, classroom observations and anecdotal records, analysis of work samples, and work inventories.

Different kinds of assessment produce different kinds of information. If a teacher needs to know whether a student can read a particular textbook, there are many sources of information available to her. She can consult districtwide achievement tests in

BEST COPY AVAILABLE

reading, estimate the level of the textbook, determine what score a student would need to read the textbook effectively, and then make a decision. However, it might be simpler for the teacher to ask the student to read a section of the text and then talk with the student about the text. This would probably be faster and more accurate than looking up test scores and conducting studies to see what kind of a test score is needed to comprehend the textbook. In general, teachers need information specific to the content and strategies they are teaching, and they can best get that information through assessments built around their daily educational tasks.

The public and policy makers have different needs from teachers. In general they need to know whether the school, school district, and state are effectively educating the students in their charge. For this purpose they need to collect information about many students and they need to know how those students stand in relation to other students across the United States or in relation to some specific standards set by the state. For these purposes, standardized norm-referenced or criterion-referenced tests are efficient and can give a broad picture of achievement on certain kinds of tasks. These kinds of tests are used most commonly for high-stakes decisions regarding schools and school districts.

BEST COPY AVAILABLE

Why does using tests for high-stakes decisions cause problems?

There are several possible problematic outcomes of high-stakes testing. These include making bad decisions, narrowing the curriculum, focusing exclusively on certain segments of students, losing instructional time, and moving decision making to central authorities and away from local personnel.

Tests are imperfect. Basing important decisions on limited and imperfect information can lead to bad decisions—decisions that can do harm to students and teachers and that sometimes have unfortunate legal and economic consequences for the schools. Decision makers reduce the chance of making a bad decision by seeking information from multiple sources. However, the information from norm-referenced and criterion-referenced tests is inexpensive to collect, easy to aggregate, and usually is highly reliable; for those reasons it is tempting to try to use this information alone to make major decisions.

Another problem is that high-stakes tests have a tendency to narrow the curriculum and inflate the importance of the test. Schools should address a broad range of student learning needs, not just the subjects or parts of subject areas covered on a particular test. As the consequences for low performance are raised, teachers feel pressured to raise scores at all costs. This means they will focus their efforts on activities that they think will improve the single important score. Time spent focusing on those activities will come from other activities in the curriculum and will consequently narrow the curriculum. Most state assessments tend to focus on reading, writing, and mathematics. Too much attention to these basic subjects will marginalize the fine arts, physical education, social studies, and the sciences.

Narrowing of the curriculum is most likely to occur in high-poverty schools that tend to have the lowest test scores. Compared to students in schools in affluent communities, students in high-poverty schools receive teaching with a greater emphasis in lower level skills, and they have limited access to instruction focusing on higher level thinking. A recent survey in one state that uses high-stakes assessments found that 75% of classroom teachers surveyed thought the state assessment had a negative impact on their teaching (Hoffman et al., in press).

Another way that educators sometimes respond to test pressure is to focus their attention on particular students. Sometimes this means that only low-performing readers get the instructional resources they need, and those doing only slightly better are ignored. Sometimes there is an attempt to raise test scores by focusing instructional initiatives on those students scoring just below cut-off points, and ignoring those both above or far below cutoff points. And sometimes schools place children in expensive special education programs they do not need, discourage particular children

from attending school on testing days, or encourage low-achieving students to drop out of school altogether, all in the name of getting higher test scores.

The loss of instructional time also is a negative result of high-stakes tests. The time for preparing for and taking tests is time taken away from basic instruction. The consequences of lost instructional time, particularly for low-performing students, are too great for information that can be gathered more efficiently.

Finally, we are concerned that instructional decision making in high-stakes testing situations is diverted from local teachers and is concentrated in a central authority far away from the school. The further decision making is removed from the local level of implementation, the less adaptive the system becomes to individual needs. High-stakes assessment shifts decisions from teachers and principals to bureaucrats and politicians and consequently may diminish the quality of educational services provided to students.

Do test scores improve when high-stakes assessment is mandated?

Test scores in the states with high-stakes assessment plans have often shown improvement. This could be because high-stakes pressure and competition leads teachers to teach reading more effectively. An alternative interpretation is that gains in test scores are the result of "teaching to the test" even when reading does not improve. Analyses of national reading scores do not show the substantial gains claimed by state reading assessments. Studies of norm-referenced tests in states with sustained patterns of growth in state skill assessments (for example, Texas and Kentucky) show no comparable patterns of gain. Although Texas showed steady improvement on state tests, its National Assessment of Educational Progress (NAEP) reading scores are not among the highest, and the scores did not show significant improvement between 1992 and 1998 (U.S. Department of Education, 1999). This may be the result of high-stakes assessments that tend to narrow the curriculum and emphasize only parts of what students need to learn to become successful readers.

Why don't we just end high-stakes assessment?

It is unlikely that states using these assessments will abandon them. Indeed, the most likely scenario is for an increasing number of states to develop and adopt similar assessment plans. Tests can be useful for making state-level educational decisions, and they provide the public with at least a partial understanding of how well schools are doing. Less positively, politicians, bureaucrats, and test publishers have discovered that they can influence classroom instruction through the use of high-stakes tests. Tests allow these outside parties to take control away from local educational authorities without assuming the responsibilities of educating the students.

BEST COPY AVAILABLE

Is there a way to help states monitor student success in the curriculum?

If the intent of state assessments is to measure how well students are learning the outcomes identified in the state curriculum framework, then one way students' success can be monitored is by following the NAEP model with selective sampling across student populations and across content areas on a systematic basis. This model monitors achievement without encouraging high-stakes testing. The tests are directed toward particular grade levels and are not given every year. A sampling procedure is used so very few students actually participate in testing. NAEP is designed to give a report card on general achievement levels in the basic subject areas over time.

Many aspects of the NAEP assessment in reading are commendable. The NAEP sampling strategy has been useful in keeping efficiency high and maintaining a focus on the questions that the national assessment is designed to address. Sampling also has provided NAEP with an opportunity to experiment with a wide variety of testing formats and conditions. Such a strategy would avoid most of the problems associated with teaching to the test. This type of plan would reflect sound principles of instructional design and assessment.

In the book *High Stakes: Testing for Tracking, Promotion, and Graduation* (Heubert & Hauser, 1999), the following basic principles for test use are presented:

• The important thing about a test is not its validity in general, but its validity when used for a specific purpose. Thus, tests that are valid for influencing classroom practice, "leading" the curriculum, or holding schools accountable are not appropriate for making high-stakes decisions about individual student mastery unless the curriculum, the teaching, and the tests are aligned.

• Tests are not perfect. Test questions are a sample of possible questions that could be asked in a given area. Moreover, a test score is not an exact measure of a student's knowledge or skills. A student's score can be expected to vary across different versions of a test—within a margin of error determined by the reliability of the test—as a function of the particular sample of questions asked and/or transitory factors, such as the student's health on the day of the test. Thus, no single test score can be considered a definitive measure of a student's knowledge.

• An educational decision that will have a major impact on a test taker should not be made solely or automatically on the basis of a single test score. Other relevant information about the student's knowledge and skills should also be taken into account.

• Neither a test score nor any other kind of information can justify a bad decision. Research shows that students are typically hurt by a simple retention and repetition of a grade in school without remedial and other instructional support services. In the absence of effective services better tests will not lead to better educational outcomes. (p. 3)

State testing programs should respect these basic principles.

What are the recommendations of the International Reading Association regarding high-stakes reading assessments?

In framing our recommendations the Association would like to stress two points. First, we recognize accountability is a necessary part of education. Concerns over high-stakes tests should not be interpreted as fear of or disregard for professional accountability. Second, the intent in this position statement is not to blame policy makers for the current dilemma with high-stakes testing.

Our recommendations begin with a consideration of teachers and their responsibility to create rich assessment environments in their classrooms and schools. Next, we suggest that researchers must continue to investigate how assessment can better serve our educational goals. Third, we stress the importance of parents and community members in bringing balance to the assessment design. Finally, we offer recommendations to policy makers for developing a plan of action.

Recommendations to teachers:

• Construct more systematic and rigorous assessments for classrooms, so that external audiences will gain confidence in the measures that are being used and their inherent value to inform decisions.

• Take responsibility to educate parents, community members, and policy makers about the forms of classroom-based assessment, used in addition to standardized tests, that can improve instruction and benefit students learning to read.

• Understand the difference between ethical and unethical practices when teaching to the test. It is ethical to familiarize students with the format of the test so they are familiar with the types of questions and responses required. Spending time on this type of instruction is helpful to all and can be supportive of the regular curriculum.

It is not ethical to devote substantial instructional time teaching to the test, and

BEST COPY AVAILABLE

it is not ethical to focus instructional time on particular students who are most likely to raise test scores while ignoring groups unlikely to improve.

• Inform parents and the public about tests and their results.

• Resist the temptation to take actions to improve test scores that are not based on the idea of teaching students to read better.

Recommendations to researchers:

• Conduct ongoing evaluations of high-stakes tests. These studies should include but not be limited to teacher use of results, impact on the curriculum focus, time in testing and test preparation, the costs of the test (both direct and hidden), parent and community communication, and effects on teacher and student motivations.

There are few data on the impact of tests on instruction. Good baseline data and follow-up studies will help in monitoring the situation. These studies should not be left to those who design, develop, and implement tests; they should be conducted by independent researchers.

• Find ways to link performance assessment alternatives to questions that external audiences must address on a regular basis. Researchers must continue to offer demonstrations of ways that data from performance assessments can be aggregated meaningfully. This strategy will allow them to build trustworthy informal assessments.

Recommendations to parents, parent groups, and child advocacy groups:

• Be vigilant regarding the costs of high-stakes tests on students. Parents must ask questions about what tests are doing to their children and their schools. They cannot simply accept the "we're just holding the school accountable" response as satisfactory. They must consider cost, time, alternative methods, and emotional impact on students as a result of these tests.

• Lobby for the development of classroom-based forms of assessment that provide useful, understandable information, improve instruction, and help children become better readers.

Recommendations to policy makers:

• Design an assessment plan that is considerate of the complexity of reading, learning to read, and the teaching of reading. A strong assessment plan is the best ally of teachers and administrators because it supports good instructional decision making and good instructional design. Consider the features of good assessment as outlined in *Standards for the Assessment of Reading and Writing* (International Reading Association & National Council of Teachers of English, 1994) in designing an assessment plan. Be aware of the pressures to use tests to make high-stakes decisions.

• When decisions about students must be made that involve high-stakes outcomes (e.g., graduation, matriculation, awards) rely on multiple measures rather than just performance on a single test. The experiences in England with high-stakes assessment have been instructive. England has moved to an assessment system that values teacher informal assessments, ongoing performance assessments, portfolios, teacher recommendations, and standardized testing. The triangulation of data sources leads to more valid decision making.

• Use sampling strategies when assessments do not involve decisions related to the performance of individual students (e.g., program evaluation). Sampling is less intrusive, less costly, and just as reliable as full-scale assessment plans. Sampling strategies also provide an opportunity to design alternate forms and types of assessments. Such a variety of assessments encourages careful inspection of issues of validity and reliability.

• Do not use incentives, resources, money, or recognition of test scores to reward or punish schools or teachers. Neither the awards (e.g., blue ribbon schools) nor the punishing labels (e.g., low-performing schools) are in the interest of students or teachers. The consequences of achieving or not achieving in schools are real enough. Well-intentioned efforts to recognize achievement often become disincentives to those who need the most help.
• Do not attempt to manipulate instruction through assessments. In other words, do not initiate, design, or implement high-stakes tests when the primary goal is to affect instructional practices. Ask the question, "Is the primary goal of the assessment to collect data that will be used to make better decisions that impact the individual students taking the test?" If the answer is "no," high-stakes tests are inappropriate.

The pattern of testing as the preferred tool to manipulate teaching continues to expand. We call on educators, policy makers, community leaders, and parents to take a common-sense look at the testing in schools today. Visit classrooms. Talk to teachers. Listen to teachers talk about the curriculum and the decisions they are making. Talk to the teachers about the kinds of assessments they use in the classroom and how they use collected data. To be opposed to large-scale, high-stakes testing is not to be opposed to assessment or accountability. It is to affirm the necessity of aligning our purposes and goals with our methods.

References

Heubert, J.P., & Hauser, R.M. (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press.

Hoffman, J., Paris, S., Patterson, E.U., Pennington, J., & Assaf, L.C. (in press). *High-stakes assessment in the language arts: The piper plays, the players dance, but who pays the price?* In J. Flood, J.M. Jensen, D. Lapp, & J. Squire (Eds.), *Handbook of research on teaching the English language arts* (2nd ed.).

International Reading Association & National Council of Teachers of English. (1994). *Standards for the assessment of reading and writing.* Newark, DE: International Reading Association; Urbana, IL: National Council of Teachers of English.

U.S. Department of Education. (1999). *The NAEP 1998 reading report card for the nation and the states* (NCES 1999-459). Washington, DC: Author.

Suggested Readings

1. What does the term *high-stakes testing* mean?

Downing, S., & Haladyna, T. (1996). A model for evaluation of high-stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practice, 5*(1), 5–12.

Popham, W. (1987). Can high-stakes tests be developed at the local level? *NASSP Bulletin, 71*(496), 77–84.

2. Why are we concerned with high-stakes testing?

Pipho, C., & Hadley, C. (1985). *State activity: Minimum competence testing as of January 1985* (Clearinghouse notes). Denver, CO: Education Commission of the States.

3. Is testing an important part of good educational design?

International Reading Association. (1995). *Reading assessment in practice.* Newark, DE: Author.

4. Why does using tests for high-stakes decisions cause problems?

Allington, R.L., & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York State. *Educational Policy, 6*(4), 396–413.

Madaus, G.F. (1985). Test scores as administrative mechanisms in educational policy. *Phi Delta Kappan, 66*(9), 611–617.

Mathison, S. (1989). *The perceived effects of standardized testing on teaching and curriculum.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

McGill-Franzen, A., & Allington, R.L. (1993). Flunk 'em or get them classified: The contamination of primary grade accountability data. *Educational Researcher, 22*(1), 19–22, 34.

Paris, S.G. (1998). Why learner-centered assessment is better than high-stakes testing. In N.M. Lambert & B.L. McCombs (Eds.), *How students learn: Reforming schools through learner-centered education* (pp. 189–209). Washington, DC: American Psychological Association.

5. Do test scores improve when high-stakes assessment is mandated?

Cornet, H.D., & Wilson, B.L. (1991). *Testing reform and rebellion.* Norwood, NJ: Ablex.

Resnick, D.P., & Resnick, L.B. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher, 14*(4), 5–20.

Wise, A.E. (1990, January 10). A look ahead: Education and the new decade. *Education Week,* p. 30.

6. Why don't we just end high-stakes assessment?

Madaus, G.F. (1985). Test scores as administrative mechanisms in educational policy. *Phi Delta Kappan, 66*(9), 611–617.

7. Is there a way to help states monitor student success in the curriculum?

Linn, R.L. (1993). Educational assessment: Expanded expectations and challenges. *Educational, Evaluation and Policy Analysis, 15*(1), 1–16.

Messick, S. (1993). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Washington, DC: American Council on Education.

Moss, P. (1998). The role of consequences in validity theory. *Educational measurement: Issues and practice, 17*(2), 6–12.

Adopted by the Board of Directors, May 1999
Board of Directors at Time of Adoption

Kathryn A. Ransom, President
Carol Minnick Santa, President-Elect
Carmelita K. Williams, Vice President
Alan E. Farstrup, Executive Director

Kathryn H. Au
Betsy M. Baker
Patricia A. Edwards
James V. Hoffman
Adria F. Klein
Diane L. Larson
John W. Logan
Lesley Mandel Morrow
Timothy Shanahan

This brochure may be purchased from the International Reading Association in quantities of 10, prepaid only. (Please contact the Association for pricing information.) Single copies are free upon request by sending a self-addressed, stamped envelope. Requests from outside the United States should include an envelope, but postage is not required.

©1999 International Reading Association
Cover and inside photo 3: Little Bighorn Photos
Inside photo 1, 2, & 4: Robert Finken

INTERNATIONAL
Reading Association
800 Barksdale Road
PO Box 8139
Newark, Delaware 19714-8139, USA
Phone: 302-731-1600
Fax: 302-731-1057
Web site: www.reading.org

99-19 pub 6/99 1035 7/99